DOCUMENT RESUME

ED 186 267 SE 030 707

AUTHOR Good, Ron
TITLE A Workshop on Explained Variance: Its Importance and Techniques for Calculation.
PUB DATE Apr 80
NOTE 16p.; Paper presented at the annual meeting of the National Association for Research in Science Teaching (53rd, Boston, MA, April 11-13, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; Error of Measurement; *Research; Research Design; Research Problems; Research Utilization; *Science Education; *Statistical Analysis; *Statistical Significance

ABSTRACT

Knowledge of the magnitude of effect(s) of an experimental study in science education should be of utmost concern to researchers in the field, but is often not reported. This document describes the concept of "explained variance" in analysis of variance designs and then explains how it can be calculated and reported. Reporting the magnitude of effect(s) of an experimental study in science education allows one to assess the practical significance of the research. Three indices for estimating magnitude of effect(s) are eta-squared, epsilon-squared, and omega-squared. In this document, the last index is used in some detail to illustrate techniques for calculating its value in various designs in science education research. (Author/MK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED186267

A WORKSHOP ON

"Explained Variance: Its Importance and Techniques for Calculation

by

Ron Good

Science Education

Florida State University

presented at the

53rd Annual Meeting of the

National Association for Research in

Science Teaching, Boston, Massachusetts,

April 12, 1980

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ron Good

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE 030 707

STATISTICAL SIGNIFICANCE

A "Mini-Test"

T F 1. Statistical significance is reported as a probability.

T F 2. A probability level of .05 is normally used to reject a null hypothesis.

T F 3. That .05 figure (alpha level) was arbitrarily set in doing agricultural research.

T F 4. Any difference in test scores of two groups can be statistically significant if the sample size is large enough.

T F 5. This type of information (statistical significance) says nothing about the relative magnitude of the effect of a treatment on a dependent variable of interest (e.g., critical thinking, creativity, etc.).

T F 6. Many experimental studies in education are unable to account for more than 2 or 3% of the variance in measures of dependent variables of interest even though they report "statistically significant" differences.

Its Importance

If you identified all of the items on the mini-test as correct statements, you probably have a reasonably accurate understanding of the limitations of the practical importance of statistical significance.

A survey of experimental studies reported in such journals as Science Education and Journal of Research in Science Teaching will show that a determination of statistical significance of differences in test scores with different groups of subjects, is assumed by the authors to be all that is necessary in order to draw conclusions about the importance of some experimental treatment.

If the probability associated with an F ratio is .05, then the difference in groups is often said to be signficant. If the probability is .01, it is seen as "very" significance; and, if the probability is .001 or less, the researcher often concludes that the experimental treatment is "extremely" significant in producing some effect. As Friedman (1968) has noted, "the basic problem is that a probability value by itself provides little information concerning the magnitude of the observed effect."

Any variable that produces a measurable effect can be a "statistically" significant factor if the sample size is large enough. This is not to minimize the importance of the concept of "power" (see Cohen, 1969 or Penick and Brewer, 1972), which is the probability of rejecting a null hypothesis

when it should be rejected. However, both statistical significance and power are probabilities which provide reliability information but do not allow for an estimate of the magnitude of the treatment effect (or other variables of interest).

When correlation measures are used for data analysis, r^2 is commonly reported as the amount of variance explained by or accounted for when comparing test scores. A correlation coefficient (r) of .50 between scores A and B indicates that 25% (r^2) of the variance in scores from one measure is explained by (accounted for) variance in scores on the other measure. It is not uncommon to see correlation coefficients reported as statistically significant at the .05 level even though they account for only 5% (or less) of the variance. Said another way, at least 95% of the variance in scores on measure A is unexplained by the variance in scores on measure B.

This concept of explained variance has been extended to analysis of variance (ANOVA), the technique most often used in experimental studies in education. "Magnitude of effect" rather than explained variance is often used to identify this technique in ANOVA designs and these terms will be used interchangeably throughout the remainder of the paper. Although Cohen (1966) and Friedman (1968) discuss applications of the technique to nonparametric statistics (chi-square) we will limit our discussion to ANOVA designs common to experimental research in science education.

Techniques for Calculation

Many experimental designs have been described for research in the social sciences. Winer (1971) and Lindman (1974) are two excellent sources on

experimental design, the latter dealing exclusively with ANOVA designs. Specific techniques for calculating explained variance (magnitude of effect) in ANOVA designs depend on whether factors are fixed, mixed, or random. Experimental studies in science education usually involve fixed factors. This means that the experimenter has deliberately chosen one or more treatments from among a theoretically infinite number. Random selection of treatments is uncommon simply because the experimenter typically wants to verify certain assumptions about specific treatments.

Although three methods of estimating explained variance in ANOVA designs have been proposed and briefly summarized by Halderson and Glasnapp (1972) and Good and Fletcher (in press), omega squared (ω^2) will be used to illustrate techniques for calculating the magnitude of main and interaction effects.

Basic Equation for Omega Squared (ω^2)

Omega squared (ω^2) was proposed by Hays (1963) as a means to estimate the magnitude of effects in ANOVA designs and is defined as a ratio of one variance component to the sum of all variance components:

$$\omega^2 = \frac{SS_A - df_A\, MS_e}{SS_T + MS_e} \qquad (1)$$

where SS_A = treatment sum of squares

df_A = degrees of freedom for treatment

MS_e = error mean square

SS_T = total sum of squares

Hays restricted the use of ω^2 to ANOVA designs where all factors are fixed, and most of this paper will deal with such designs. However, Fleiss (1969) and Vaughn and Corballis (1969) have expanded the application of ω^2 to mixed and random ANOVA designs.

Applications of the Basic Equation (1)

Case 1: JRST, 17, 75-81.

The first study that will be used to illustrate the application of equation (1) analyzed the effect of student use of manipulative models on chemistry achievement. A 2 x 3 nonrepeated, fixed factor design was used and both main effects (treatment and cognitive level) were found to be statistically significant beyond the .01 level. Although we know that there is little probability that the differences in test scores between treatment groups would occur due to chance alone, very little is known about the magnitude of the effect of experimental treatment.

ω^2 for treatment

$SS_A = 277.37$

$df_A = 1$

$MS_e = 30.82$

$SS_T = 7425.00$

$$\omega^2 = \frac{277.37 - (1)\ 30.82}{7425 + 30.82} = \frac{246.55}{7455.82} = .03$$

This means that 3% of the variance in test scores can be accounted for by the treatment.

ω^2 for cognitive groups

$SS_A = 443.65$

$df_A = 2$

$MS_e = 30.82$

$SS_T = 7425.00$

$$\omega^2 = \frac{443.65 - (2)\ 30.82}{7425 + 30.82} = \frac{382.01}{7455.82} = .05$$

This means that 5% of the variance in test scores can be accounted for by the level of cognitive development.

ω^2 for interaction (treatment x cognitive groups)

With the fixed effects design, interaction as well as main effects can be used in ω^2 calculations. Although the probability associated with the F ratio for interaction effects in this JRST study is considerably larger than the "magic" .05 level (.213 in this case), ω^2 is calculated here to determine the total "explained variance" that can be attributed to this study.

$SS_A = 95.98$

$df_A = 2$

$MS_e = 30.82$

$SS_T = 7425.00$

$$\omega^2 = \frac{95.98 - (2)\ 30.82}{7425 + 30.82} = \frac{34.34}{7455.82} = .01$$

This means that 1% of the variance in test scores can be accounted for by the treatment x cognitive level interaction.

For this particular study reported in JRST, a total of 9% of the variance in the dependent variable scores (chemistry test) can be accounted for by all effects, main and interaction. Said another way, 91% of the variance in the dependent variable scores must be due to factors other than those analyzed in the design. The treatment effect accounted for 3% of the variance in the chemistry scores. This figure is not much less than the variance accounted for by main effects in most studies reported in science education journals, even though statistically significant p values ($p \leq .05$) are listed by the authors.

Case 2: Science Education, 63, 37-44

This study used a 2 x 3 nonrepeated, fixed factor ANOVA design. The treatment consisted of two readability levels of the same science content administered to three groups of students with high, moderate, and low reading ability. The F ratio for the treatment was 1.97 ($p > .05$) and for the reading abilities it was 62.39 ($p < .01$).

ω^2 for reading ability

$SS_A = 1997.07$

$df_A = 2$

$MS_e = 16.00$

$SS_T = 3765.13$

$$\omega^2 = \frac{1997.07 - (2)\ 16}{3765.13 + 30.82} = \frac{1965.07}{3795.95} = .52$$

Thus, 52% of the variance in science achievement test scores was accounted for by reading ability of the students.

ω^2 for treatment

$SS_A = 31.58$

$df_A = 1$

$MS_e = 16$

$SS_T = 3765.13$

$$\omega^2 = \frac{31.58 - (1)\ 16}{3765.13 + 16} = \frac{15.58}{3781.13} = \underline{.00}$$

Thus, the treatment in this study accounted for less than 1% of the variance in science achievement test scores.

Using F Ratios to Calculate ω^2

Since an F ratio is merely a ratio of mean squares, ω^2 can be calculated using an equation suggested by Lindman (1974, p. 54):

$$\omega^2 = \frac{F - 1}{F - 1 + N/(I-1)} \quad (2)$$

Where N = sample size

I = number of levels or groups

For Case 2 the F ratio for reading abilities was 62.39, N was

114 and I was 3. Using these values in equation (2) we get:

$$\omega^2 = \frac{62.39 - 1}{62.39 - 1 + 114/3 - 1} = \frac{61.39}{118.39} = \underline{.52}$$

It should be emphasized again that both equation (1) and equation (2) are applicable to fixed effects only.

Case 3: JRST, 14, 105-115

In analysis of covariance designs, metric independent variables are used in conjunction with nonmetric independent variables (or factors). The covariate is a metric independent variable and in this study the amount of time spent in an audio-tutorial center served as the covariate. A 2 x 3 factorial design was used to determine the effects of "kinetic structure" of audio tapes and student verbal ability on achievement and on total time spent (covariate). Both main effects (kinetic structure and verbal ability) were statistically significant.

ω^2 for treatment (kinetic structure)

$SS_A = 213.800$

$df_A = 1$

$MS_e = 45.427$

$SS_T = 6020.714$

$$\omega^2 = \frac{213.800 - (1)\ 45.427}{6020.714 + 45.427} = \frac{168.373}{} = \underline{.03}$$

Thus, the magnitude of the treatment effect accounted for 3% of the total test score variance.

ω^2 for verbal ability

$SS_A = 453.539$

$df_A = 2$

$MS_e = 45.427$

$SS_T = 6020.714$

$$\omega^2 = \frac{453.539 - (2)\ 45.427}{6020.714 + 45.427} = \frac{362.685}{} = .06$$

The magnitude of the effect of verbal ability of students amounted to 6%, or twice that of the treatment effect.

ω^2 In More Complicated Designs

The large majority of designs used in experimental studies in science education research fall within the category, nonrepeated, fixed factor. Either equation (1) or equation (2) will provide an estimate of the magnitude of main or interaction effects with this type of design. When random and/or repeated factors are a part of the design, modifications in the basic formula are needed. Equations using sum of squares (SS) and mean square (MS) information are presented in a paper by Halderson and Glasnapp (1972) while equations using F ratios can be found in a paper by Fleiss (1969).

In the repeated measures design, each subject receives each treatment. The treatments are the levels of the fixed factor and the subjects are the levels of the random factor. The mean square (MS) of the random factor

is often referred to as "between" subjects mean square and the interaction mean square as the "within" subjects mean square.

In general, the fixed effects design is more robust than mixed or random designs in terms of underlying assumptions. Lindman (1974, p. 133) discusses the problem of robustness in random and mixed designs.

Recommendations

The main recommendation, of course, is that experimental researchers in science education should calculate the magnitude of an effect in addition to determining its statistical significance. This seems so obvious that it is difficult to understand why it is not done with regularity in all experimental research in education.

The second recommendation is less easy to defend but appears to be just as necessary in explained variance calculations as is an alpha level of .05 in statistical significance calculations. The recommendation is that science educators establish categories of small, medium and large effect sizes for research involving experimental curriculum and instruction variables. A recent survey by Good and Fletcher (in press) of 16 studies reported in JRST, found main effect sizes (ω^2) for instructional variables, ranging from under 3% to over 30%. Based on this survey and on a consideration of the size of an effect required for making decisions about curriculum and instruction in science education, the following categories are offered as a place to begin the process:

Effect Size	Small	Medium	Large
ANOVA	0-.10	.11-.25	.26+

It is recognized that as one goes from the simple one way, fixed effects ANOVA design to the more complex n-way random and mixed designs, effect size decisions become more involved. Also, certain types of treatments are more likely to have an effect on certain dependent variables and, so, one would expect higher values for explained variance. In spite of these problems, researchers in science education should move in the direction of establishing guidelines that will assist those who translate theory and research into school practice. Communicating the relative magnitude of an experimental effect is one important way of doing this.

References

Cohen, Jacob. "Some Statistical Issues in Psychological Research." In B.B. Wolman (Ed.) Handbook of Clinical Psychology. New York: McGraw-Hill, 1965.

Cohen, Jacob. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977.

Fleiss, Joseph. "Estimating the Magnitude of Experimental Effects." Psychological Bulletin, 72, 273-276, 1969.

Friedman, Herbert. "Magnitude of Experimental Effect and a Table for Its Rapid Estimation." Psychological Bulletin, 70, 245-251, 1968.

Good [illegible] and Harold Fletcher. "Reporting Explained Variance," Journal of Research in Science Teaching, in press.

Halderson, Judith and Douglas Glasnapp. "Generalized Rules for Calculating the Magnitude of an Effect in Factorial and Repeated Measures ANOVA Designs." American Educational Research Journal, 9, 301-310, 1972.

Hays, W.L. Statistics for Psychologists. New York: Holt, Rinehart, & Winston, 1963.

Lindman, Harold. Analysis of Variance in Complex Experimental Designs. San Francisco: W. H. Freeman, 1974.

Penick, John and James Brewer, "The Power of Statistical Tests in Science Teaching Research." Journal of Research in Science Teaching, 9, 377-381, 1972.

Vaughan, Graham and Michael Corballis. "Beyond Tests of Significance: Estimating Strength of Effects in Selected ANOVA Designs." Psychological Bulletin, 72, 204-213, 1969.

Winer, B.J. Statistical Principles in Experimental Design. New York: Mc Graw-Hill, 1971.

RG/olt/Th-1